

Published in final edited form as:

Science. 2011 May 20; 332(6032): 960–963. doi:10.1126/science.1203810.

The compact *Selaginella* genome identifies changes in gene content associated with the evolution of vascular plants

Jo Ann Banks¹, Tomoaki Nishiyama^{2,3}, Mitsuyasu Hasebe^{3,4,5}, John L. Bowman^{6,7}, Michael Gribskov⁸, Claude dePamphilis⁹, Victor A. Albert¹⁰, Naoki Aono⁴, Tsuyoshi Aoyama^{4,5}, Barbara A. Ambrose¹¹, Neil W. Ashton¹², Michael J. Axtell⁹, Elizabeth Barker¹², Michael S. Barker¹³, Jeffrey L. Bennetzen¹⁴, Nicholas D. Bonawitz¹⁵, Clint Chapple¹⁵, Chaoyang Cheng³, Luiz Gustavo Guedes Correa¹⁶, Michael Dacre¹⁷, Jeremy DeBarry¹⁴, Ingo Dreyer¹⁸, Marek Elias¹⁹, Eric M. Engstrom²⁰, Mark Estelle²¹, Liang Feng²², Cédric Finet²³, Sandra K. Floyd⁶, Wolf B. Frommer²⁴, Tomomichi Fujita²⁵, Lydia Gramzow²⁶, Michael Gutensohn²⁷, Jesper Harholt²⁸, Mitsuru Hattori^{29,30}, Alexander Heyl³¹, Tadayoshi Hirai^{3,32}, Yuji Hiwatashi^{4,5}, Masaki Ishikawa³, Mineko Iwata³, Kenneth G. Karol³³, Barbara Koehler¹⁶, Uener Kolukisaoglu³⁴, Minoru Kubo³, Tetsuya Kurata^{3,35}, Sylvie Lalonde²⁴, Kejie Li⁸, Ying Li^{8,36}, Amy Litt³³, Eric Lyons³⁷, Gerard Manning¹⁷, Takeshi Maruyama³⁸, Todd P. Michael³⁹, Koji Mikami⁴⁰, Saori Miyazaki^{4,41}, Shin-ichi Morinaga^{4,42}, Takashi Murata^{4,5}, Bernd Mueller-Roeber¹⁶, David R. Nelson⁴³, Mari Obara^{3,44}, Yasuko Oguri³, Richard G. Olmstead⁴⁵, Naoko Onodera^{3,46}, Bent Larsen Petersen²⁸, Birgit Pils⁴⁷, Michael Prigge²¹, Stefan A. Rensing^{48,49,50}, Diego Mauricio Riaño-Pachón^{51,52}, Alison W. Roberts⁵³, Yoshikatsu Sato³, Henrik Vibe Scheller^{37,54}, Burkhard Schulz²⁷, Christian Schulz⁵⁵, Eugene V. Shakhov⁵⁶, Nakako Shibagaki⁵⁷, Naoki Shinohara³, Dorothy E. Shippen⁵⁶, Iben Sørensen²⁸, Ryo Sotooka⁵⁸, Nagisa Sugimoto³, Mamoru Sugita²⁹, Naomi Sumikawa⁴, Milos Tanurdzic⁵⁹, Günter Theißen²⁶, Peter Ulvskov²⁸, Sachiko Wakazuki³, Jing-Ke Weng^{15,60}, William W.G.T. Willats²⁸, Daniel Wipf⁶¹, Paul G. Wolf⁶², Lixing Yang¹⁴, Andreas D. Zimmer^{48,63}, Qihui Zhu¹⁴, Therese Mitros⁶⁴, Uffe Hellsten⁶⁵, Dominique Loqué⁵⁴, Robert Otillar⁶⁵, Asaf Salamov⁶⁵, Jeremy Schmutz⁶⁵, Harris Shapiro⁶⁵, Erika Lindquist⁶⁵, Susan Lucas⁶⁵, Daniel Rokhsar^{64,65}, and Igor V. Grigoriev⁶⁵

¹Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA

²Advanced Science Research Center, Kanazawa University, Kanazawa, 920-0934 Japan

³ERATO, Japan Science and Technology Agency, Okazaki 444-8585, Japan ⁴National Institute

for Basic Biology, Okazaki 444-8585, Japan ⁵Department of Basic Biology, School of Life

Science, The Graduate University for Advanced Studies, Okazaki 444-8585, Japan ⁶School of

Biological Sciences, Monash University, Clayton Campus, Melbourne, Victoria 3800, Australia

⁷Section Plant Biology, University of California, Davis, CA 95616, USA ⁸Department of Biological

Sciences, Purdue University, West Lafayette, IN 47907, USA ⁹Department of Biology and Huck

Institutes of Life Sciences, Pennsylvania State University, University Park, PA 16802, USA

¹⁰Department of Biological Sciences, University at Buffalo, Buffalo, New York 14260 ¹¹The New

York Botanical Garden, Bronx, NY 10458, USA ¹²Department of Biology, University of Regina,

3737 Wascana Parkway, Regina, SK, S4S 0A2, Canada ¹³Department of Ecology & Evolutionary

Biology, University of Arizona, Tucson, AZ 85721, USA ¹⁴Department of Genetics, University of

Georgia, Athens, GA 30602, USA ¹⁵Department of Biochemistry, Purdue University, West

Lafayette, IN 47907, USA ¹⁶Department of Molecular Biology, University of Potsdam, Potsdam-

Golm 14476, Germany ¹⁷Razavi Newman Center for Bioinformatics, Salk Institute for Biological

Studies 10010 North Torrey Pines Rd, La Jolla, CA, USA ¹⁸Heisenberg-Group BPMPB, University

of Potsdam, 14476 Potsdam-Golm, Germany ¹⁹Faculty of Science, Department of Botany,

Charles University in Prague, 128 01 Prague, Benatska 2, Czech Republic ²⁰Department of

Biology, The College of William and Mary, Williamsburg, VA 23187, USA ²¹Section of Cell and

Developmental Biology, University of California, San Diego, La Jolla, CA 92093, USA ²²Institute

of Bioinformatics, University of Georgia, Athens, GA 30602, USA ²³Laboratoire de Reproduction et Développement des Plantes, Ecole Normale Supérieure de Lyon, Lyon F-69364, France ²⁴Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA ²⁵Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan ²⁶Department of Genetics, Friedrich Schiller University Jena, D-07743, Jena, Germany ²⁷Department of Horticulture and Landscape Architecture, Purdue University, W. Lafayette, IN 47907, USA ²⁸Department of Plant Biology and Biotechnology, University of Copenhagen, Frederiksberg C 1871, Denmark ²⁹Center for Gene Research, Nagoya University, Nagoya 464-8602, Japan ³⁰Department of Chemistry, School of Science, The University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan ³¹Institute for Biology/Applied Genetics/DCPS, Freie Universität Berlin, 14195 Berlin, Germany ³²Graduate School of Life and Environmental Sciences, Gene Research Center, University of Tsukuba, Tsukuba 305-8572, Japan ³³The New York Botanical Garden, Bronx, NY 10458-5126, USA ³⁴Center for Life Science Automation, University of Rostock, D-18119 Rostock, Germany ³⁵Plant Global Education Project, Graduate School of Biological Sciences, Nara Institute of Science and Technology, Nara 630-0192, Japan ³⁶Division of Biomedical Statistics and Informatics, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA ³⁷Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA ³⁸Graduate School of Life Science, Hokkaido University, Sapporo 060-0810, Japan ³⁹Department of Plant Biology and Pathology, Rutgers, The State University of New Jersey and Waksman Institute of Microbiology, 190 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA ⁴⁰Faculty of Fisheries Sciences, Hokkaido University, Hakodate 041-8611, Japan ⁴¹National Institute of Genetics, Mishima 411-8540 Japan ⁴²Graduate School of Arts and Sciences, The University of Tokyo, Tokyo 153-8902, Japan ⁴³Department of Microbiology, Immunology and Biochemistry, University of Tennessee, Memphis, TN 38163, USA ⁴⁴Japan Science and Technology Agency, Sapporo 060-0819, Japan ⁴⁵Department of Biology, University of Washington, Seattle, WA 98195-5325, USA ⁴⁶Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada ⁴⁷Hamilton Bonaduz AG, Bonaduz 7402, Switzerland ⁴⁸Faculty of Biology, University of Freiburg, Freiburg 79104, Germany ⁴⁹BIOSS Centre for Biological Signaling Studies, University of Freiburg, Freiburg 79104, Germany ⁵⁰FRISYS Freiburg Initiative for Systems Biology, University of Freiburg, Freiburg 79104, Germany ⁵¹GabiPD team, Bioinformatics Group, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm 14476, Germany ⁵²Group of Computational and Evolutionary Biology, Biological Sciences Department, Universidad de los Andes, Bogotá D.C., Colombia ⁵³Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA ⁵⁴Joint BioEnergy Institute, Feedstocks Division, Emeryville, CA 94608, USA ⁵⁵Department of Evolution and Biodiversity of Plants, Ruhr-University Bochum, Bochum 44780, Germany ⁵⁶Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA ⁵⁷Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan ⁵⁸School of Science, Hokkaido University, Sapporo 060-0810, Japan ⁵⁹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA ⁶⁰The Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA ⁶¹UMR INRA 1088/CNRS 5184, Université Bourgogne Plant-Microbe-Environment, Dijon 21065, France ⁶²Department of Biology, Utah State University, Logan, UT 84322-5305, USA ⁶³Plant Biotechnology, University of Freiburg, Freiburg 79104, Germany ⁶⁴Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA ⁶⁵US Department of Energy Joint Genome Institute, Walnut Creek Ca 94598, USA

Abstract

Vascular plants appeared ~410 million years ago then diverged into several lineages of which only two survive: the euphyllophytes (ferns and seed plants) and the lycophytes (1). We report here the genome sequence of the lycophyte *Selaginella moellendorffii* (*Selaginella*), the first non-seed

vascular plant genome reported. By comparing gene content in evolutionary diverse taxa, we found that the transition from a gametophyte- to sporophyte-dominated life cycle required far fewer new genes than the transition from a non-seed vascular to a flowering plant, while secondary metabolic genes expanded extensively and in parallel in the lycophyte and angiosperm lineages. *Selaginella* differs in post-transcriptional gene regulation, including small RNA regulation of repetitive elements, an absence of the tasiRNA pathway and extensive RNA editing of organellar genes.

Selaginella moellendorffii, like all lycophytes, has features typical of vascular plants, including a dominant and complex sporophyte generation (Fig. 1A and B) having vascular tissues with lignified cell types. Lycophytes also share traits with non-seed plants, most notably the release of haploid spores (Fig. 1C) from the sporophyte and a gametophyte generation that develops independently of the sporophyte. Because the lycophytes are an ancient lineage that diverged shortly after land plants evolved vascular tissues (Fig. 2A) (1), we sequenced the *Selaginella* genome to provide a resource for identifying genes that may have been important in the early evolution of developmental and metabolic processes unique to vascular plants.

The *Selaginella* genome was sequenced using whole-genome shotgun sequencing (2). The assembled genome size (212.6 Mbp) is twice that determined by flow cytometry (3), indicating that the assembled genome includes two haplotypes of ~106 Mbp that are 98.5% identical at the nucleotide level. A deduced haplotype has 22,285 predicted protein-coding genes, of which 37% are supported by EST sequences, and 58 microRNA (miRNA) loci (2, 4). The *Selaginella* genome lacks evidence of an ancient whole genome duplication or polyploidy (2), unlike all other sequenced land plant genomes (5–7). Gene density in *Selaginella* and *Arabidopsis*, which has a slightly larger genome size, is very similar (2), and both genomes having gene-poor regions rich in transposable elements (TEs) and other repetitive sequences (2). While fewer genes and smaller introns (2) contribute to a genome size smaller than *Arabidopsis*, this is offset by a greater proportion of TEs in *Selaginella* (37.5% vs. 15% in *Arabidopsis*)(2). LTR retrotransposons are the most abundant TEs, occupying one-third of the *Selaginella* genome (2).

Plant TEs and *MIRNA* loci are significant sources of small RNAs (sRNAs) that function to epigenetically regulate TE and gene activity (8). Several observations suggest that some aspects of epigenetic or post-transcriptional gene regulation in *Selaginella* are unique among plants. For one, the proportion of sRNAs 23–24 nt in length is extraordinarily small in the *Selaginella* sRNA population (2) compared to angiosperms (9). Nearly three-quarters of the *Selaginella* sRNAs (4) map to *MIRNA* loci and are predominantly 21 nt in length (2). In angiosperms, 24 nt siRNAs, which are generated primarily from TEs, function to silence TE activity through the RdDM pathway (10–12) and accumulate massively in specific cells of the female gametophyte (13). Since the *Selaginella* sRNA population was generated from sporophytic tissues, the 24 nt siRNA pathway may only be deployed during gametophyte development in *Selaginella*. A second distinction is the absence of *DCL4*, *RDR6* and *MIR390* loci in *Selaginella*, which are required for the biogenesis of trans-acting siRNAs (tasiRNAs) in angiosperms (2). Their absence suggests that tasiRNA-regulated processes in angiosperms, including leaf polarity (14) and developmental phase changes in the sporophyte (15, 16), are regulated differently in *Selaginella*, and possibly reflects the independent origins of foliar organs in the lycophyte and angiosperm lineages (17, 18). Finally, the *Selaginella* plastome sequence reveals an extraordinarily large number of RNA edited sites (2), as do other lycophyte organellar genomes (19, 20). This coincides with an extraordinarily large number of PPR genes in *Selaginella* (>800; (2), some of which guide RNA editing events in angiosperms (21).

Because *Selaginella* is a member of a vascular plant lineage that is sister to the euphyllophytes, we used comparative and phylogenetic approaches to identify gene origins and expansions coinciding with evolutionary innovations and losses in land plants. To identify such genes without regard to function, we compared the proteomes of the green alga *Chlamydomonas*, the moss *Physcomitrella*, *Selaginella*, and 15 angiosperm species, identified gene families that are related by homology by hierarchical clustering (2), and then mapped them onto a phylogenetic tree (Fig. 2B). The 3814 families with gene members present in all plant lineages define the minimum set of genes that were likely to be present in the common ancestor of all green plants and their descendants and include genes essential for plant function. The transition from single-celled green algae to multicellular land plant approximately doubled the gene number with the acquisition of 3006 new genes. The transition from non-vascular to vascular plant is associated with a gain of far fewer new genes (516) than the transition from a basal vascular plant to a basal euphyllophyte whose descendants include the angiosperms (1350). These numbers show that the evolution of traits unique to euphyllophytes or angiosperms required the evolution of about three times more new genes than the transition from a plant having a dominant gametophyte and simple, leafless and non-vascularized sporophyte (typified by modern bryophytes) to a plant with a dominant, vascularized and branched sporophyte with leaves.

In a second approach, we analyzed the phylogenies of genes known to function in *Arabidopsis* development (2). We identified 424 monophyletic groups of developmental genes, each group containing putatively all genes descended from a common land plant ancestral gene (Table S6). *Selaginella* and *Physcomitrella* genes are present in 377 (88%) and 356 (84%) of the 424 land plant orthologous gene groups, respectively, indicating that the common ancestor of land plants had most of the gene families known to direct angiosperm development. Conspicuous expansions of families within different lineages resulted in different numbers of land plant orthologs in each genome (Table S6). The 27 vascular plant-specific orthologous groups likely represent genes associated with developmental innovations of vascular plants. Among them are genes regulating the meristem (*CLV1* and *CLV2*), hormone signaling (*GID1*, *CTR1*) and flowering (*TFL2* and *UFO*). Interestingly, homologs of genes involved in the specification of xylem (*NST* and *VND*) (22) and phloem (*APL*) (23) in *Arabidopsis* are present in *Physcomitrella* and *Selaginella*, suggesting that the developmental programs for patterning and differentiation of vascular tissues were either present in, or co-opted from preexisting genetic programs in the ancestral land plant. The 43 groups lacking genes from *Physcomitrella* and *Selaginella* (Table S6) likely identify genes that were necessary for euphyllophyte or angiosperm developmental innovations. Among this group are genes that regulate light signaling (*FAR1*, *MIF1*, *OBP3* and *PKS1*), shoot meristem development (*AS2* and *ULT1*), hormone signaling and biosynthesis (*BRI1*, *BSU1*, *ARF16*, *ACS* and *ACO*), and flowering (*HUA1*, *EMF1*, *FT*, *TFL1* and *FD*). Altogether, these results suggest that the evolutionary transitions from a non-vascular plant to a vascular to angiosperm included the stepwise addition of components of some developmental pathways, especially those regulating meristem and hormone biology, as previously noted for the gibberellin signaling pathway (24, 25).

Genes involved in secondary metabolism were also investigated because plants synthesize numerous secondary metabolites that they use to interact with their environment. Three gene families involved in their biosynthesis, including those encoding cytochrome P450-dependent monooxygenases (P450s), BAHD acyltransferases (BAHDs) and terpene synthases (TSs), were analyzed. The largest of these in *Selaginella* is the P450 family, accounting for 1% of its predicted proteome (Table S7) (2). All three families show similar evolutionary trends, with the inferred ancestral vascular plant having a small number of genes that radiated extensively but independently within the lycophyte and angiosperm lineages (Fig. S6–13). *BAHD* and *TS* genes, which are known to be involved in the

biosynthesis of volatile odorants, are apparent only in seed plants (Figs. S12–13), likely reflecting the co-evolution of seed plants with animals that pollinate flowers or disperse seeds. The independent diversification of these gene families plus the large number of *Selaginella* genes suggest that *Selaginella* not only has the potential to synthesize a repertoire of secondary metabolites that rivals the angiosperms in complexity, but that many of them are likely to be unique. Some have been shown to be of pharmaceutical value (e.g., 26).

We have used the compact *Selaginella* genome sequence to uncover genes associated with major evolutionary transitions in land plants. Understanding their functions in *Selaginella* and other taxa, as well as acquiring the genome sequences of other informative taxa, especially charophytes, ferns and gymnosperms, will be key to understanding the evolution of plant form and function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Selaginella sequences were deposited at GenBank, accession nos. GL377566–GL378322.1, and HM173080. Genome sequencing and analysis was performed by the U.S. Department of Energy, Joint Genome Institute supported by the Office of Science of the U.S. DOE, Contract DE-AC02-05CH11231. (IG, DR, RO, AS, JS, HS, EL, SL, TM). Support provided by: NSF (JAB, JLB, CC, CD, ME, SKF, RGO, MP, CS, PGW, JW); DOE (WBF, HVS); NIH (MJA, MSB, MD, KGK, GM, DES); USDA-NIFA NRICG (AR); Lewis B and Dorothy Cullman Program (KGK, AL and BAA); Life Sciences Research Institute (NDB); Georgia Research Alliance (QZ, JLB); Busch Biomedical (TPM); Australian Research Council (JLB); Villum Kann Rasmussen Foundation and Danish Research Council (IS, JH, PU, BLP, WGTW); DFG (ID, SAF, MG, ADZ); BMBF FRISYS 0313921 (SAF); Bundesministerium fuer Bildung und Forschung (BM, DRM); Burgundy Research Council (CRB); Czech Ministry of Education (ME); NSERC (EIB, NWA, MSB); MEXT, Japan (MH, TN, TF) and JSPS (MH, TN, TF, KM, TM, MS). Dave Hurley and Steven Hentel provided computational assistance.

References and Notes

1. Kenrick P, Crane PR. *Nature*. 1997; 389:33.
2. Details given in Supplemental Online Materials.
3. Wang W, et al. *BMC Plant Biol*. 2005; 5:10. [PubMed: 15955246]
4. Axtell MJ, Snyder JA, Bartel DP. *Plant Cell*. 2007; 19:1750. [PubMed: 17601824]
5. Barker MS, Vogel H, Schranz ME. *Genome Biology and Evolution*. 2009; 1:391. [PubMed: 20333207]
6. Tang H, et al. *Science*. 2008; 320:486. [PubMed: 18436778]
7. Tang H, et al. *Genome Res*. 2008; 18:1944. [PubMed: 18832442]
8. Ghildiyal M, Zamore PD. *Nat Rev Genet*. 2009; 10:94. [PubMed: 19148191]
9. Chen D, et al. *Bioinformatics*. 2010; 26:1391. [PubMed: 20378553]
10. Kasschau KD, et al. *PLoS Biol*. 2007; 5:e57. [PubMed: 17298187]
11. Nobuta K, et al. *Nat Biotechnol*. 2007; 25:473. [PubMed: 17351617]
12. Cho SH, et al. *PLoS Genet*. 2008; 4:e1000314. [PubMed: 19096705]
13. Bourc'his D, Voinnet O. *Science*. 2010; 330:617. [PubMed: 21030645]
14. Chitwood DH, et al. *Genes Dev*. 2009; 23:549. [PubMed: 19270155]
15. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS. *Genes Dev*. 2004; 18:2368. [PubMed: 15466488]
16. Xie Z, Allen E, Wilken A, Carrington JC. *Proc Natl Acad Sci U S A*. 2005; 102:12984. [PubMed: 16129836]
17. Floyd SK, Bowman JL. *Curr Biol*. 2006; 16:1911. [PubMed: 17027487]

18. Harrison CJ, et al. *Nature*. 2005; 434:509. [PubMed: 15791256]
19. Tsuji S, et al. *J Plant Res*. 2007; 120:281. [PubMed: 17297557]
20. Grewe F, et al. *Nucleic Acids Res*. 2010
21. Knoop V. *Cell Mol Life Sci*. 2010
22. Demura T, Fukuda H. *Trends Plant Sci*. 2007; 12:64. [PubMed: 17224301]
23. Bonke M, Thitamadee S, Mahonen AP, Hauser MT, Helariutta Y. *Nature*. 2003; 426:181. [PubMed: 14614507]
24. Hirano K, et al. *Plant Cell*. 2007; 19:3058. [PubMed: 17965273]
25. Yasumura Y, Crumpton-Taylor M, Fuentes S, Harberd NP. *Curr Biol*. 2007; 17:1225. [PubMed: 17627823]
26. Cao Y, et al. *Fitoterapia*. 2009; 81:253. [PubMed: 19775597]



Figure 1. Selaginella morphology. (A) The diploid sporophyte body. Bar, 10 mm. (B) A shoot with two ranks of microphylls ("leaves") and strobili. Each microphyll of a strobilus has either a mega- or microsporangium where mega- or microspores are produced. Bar, 2 mm. (C) An orange microspore on top of a dark megaspore. These single celled haploid spores represent the beginning of the independent haploid gametophyte generation. The microgametophyte produces motile sperm and the megagametophyte eggs. Bar, 0.1 mm.

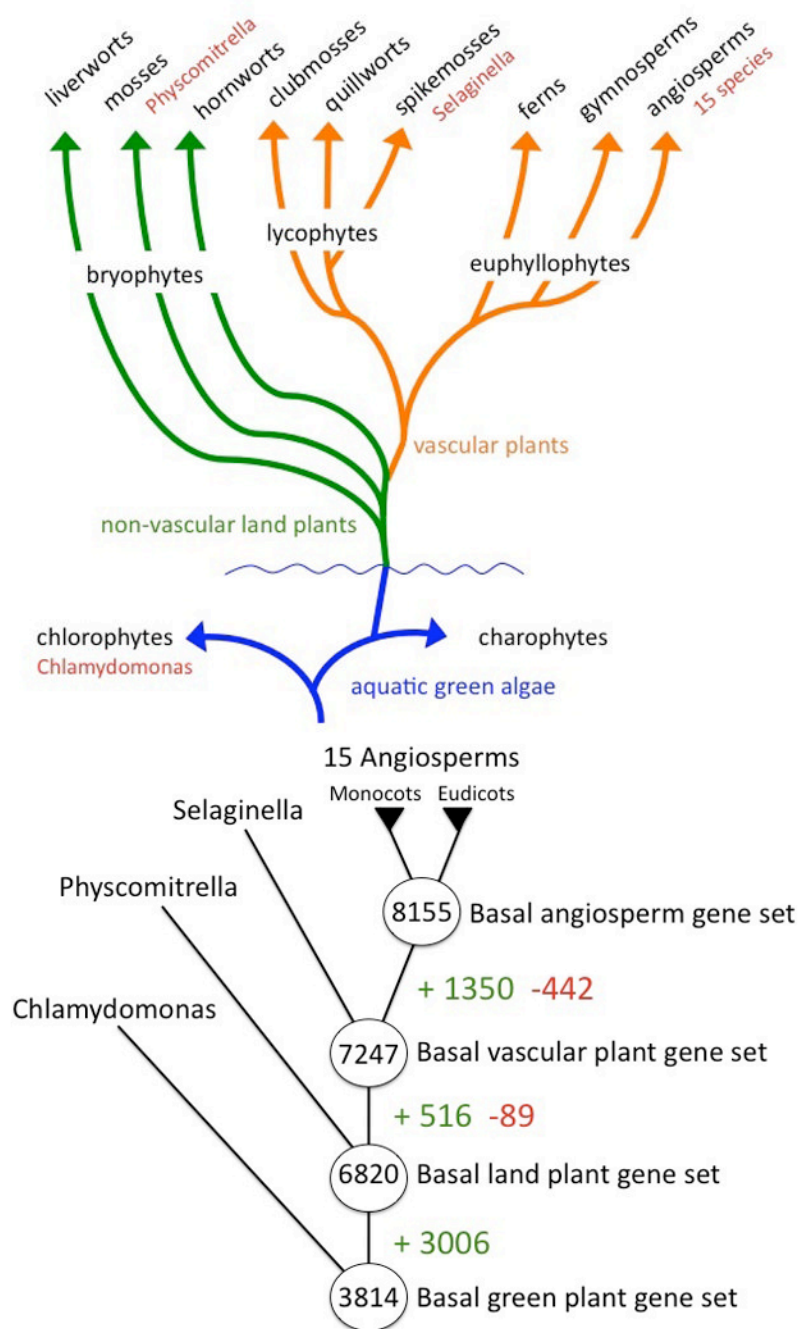


Figure 2. (A) Phylogeny of plants. Taxa in red have sequenced genomes. (B) Gene family gains (+) and losses (-) mapped onto the plant phylogenetic tree. The minimum numbers of gene families present in the ancestors of different plant lineages are circled.